Allocating Impact Evaluation Resources: Using Risk Analysis to get the Biggest Bang for your Buck[1]

Jennifer Meissner, New York State Energy Research & Development Authority, Albany, NY
Cherie Gregoire, New York State Energy Research & Development Authority, Albany, NY
Steven Meyers, Rational Energy Network, Austin, TX
Lori Megdal, Megdal & Associates, Acton, MA
Kathryn Parlin, West Hill Energy & Computing, Chelsea, VT

ABSTRACT

Gross and net program energy and demand savings are important evaluation metrics across jurisdictions and organizations. With the latest round of **New York Energy $mart^{SM}** funding, the New York State Energy Research and Development Authority (NYSERDA) is reassessing impact evaluation budgets to target areas where this work can help alleviate uncertainty and provide the highest value. NYSERDA and a contracted evaluation team led by Megdal & Associates used a unique risk analysis procedure as an objective basis for identifying and quantifying uncertainty within prior evaluation estimates of gross and net savings. The risk analysis results were one input used to decide on impact evaluation projects and funding levels.

The risk analysis developed for NYSERDA is different from others conducted in this area in that it identifies and quantifies risk based on evaluation assumptions and research design along with uncertainty from program or market unknowns. This analysis requires the development of quantitative estimates of risk from a qualitative assessment of prior research methods and findings, and is well suited as a mechanism for the continuous improvement of the reliability of evaluation estimates. Continued progress in this area holds even greater promise for exploring and refining methods to achieve more reliable evaluation results. This paper describes NYSERDA's approach to risk analysis including the process and analytic techniques, results, lessons learned, and utility of the outcomes.

Introduction: Context and NYSERDA Project Objectives

The New York State Energy Research and Development Authority (NYSERDA) operates **New York Energy $mart^{SM}**, a portfolio of System Benefit Charge (SBC)-funded energy efficiency and demand management programs in New York. NYSERDA receives approximately $175 million per year for commercial, industrial, residential, low-income and research and development efforts to increase energy efficiency, lower energy demand, and develop renewable energy technologies in New York.

Evaluations of SBC-funded programs are conducted on a budget of 2% of program cost. About one-third of this budget is allocated to impact evaluation, which includes determination of gross and net energy and demand savings. Given these budget limitations, NYSERDA must carefully consider what programs to evaluate and how. The formal risk analysis described in this paper was used for the first time in 2007 as a tool to deliberately and consistently examine impact evaluation needs and set priorities (NYSERDA).

Risk analysis has been used in the energy efficiency field to examine uncertainty in achieving energy and demand reduction targets. Typically, risk analysis examines the level and distribution of uncertainty for the expected values of program participation and savings per unit. Implementers then use this information

---

[1] The views expressed in this paper are those of the authors and do not necessarily reflect the views of the New York State Energy Research and Development Authority.

to better manage their program efforts and investments. (Hall, Jacobs & Kromer 2006; Ridge et al. 2007; Violette & Kooney 2003)

NYSERDA put a different spin on risk analysis by assessing uncertainty in the ability of the impact evaluation to develop a reliable and accurate estimate of the actual savings. NYSERDA's analysis focused on the methods used in prior evaluations, their sample sizes, and reliability issues addressed by the evaluation. This necessitates creating several quantitative indices for reliability components for both the ex-post gross and net-to-gross evaluations. Identifying and sizing up reliability issues is a significant advancement in the use of risk analysis as part of allocating evaluation budgets.

The NYSERDA risk analysis fits into the broader scope of its impact evaluation and serves four primary objectives:

- Allow NYSERDA to develop first-hand experience with this technique to better understand its potential utility to the **New York Energy $mart**^{SM} Program evaluation planning process.
- Aid in the allocation of evaluation budgets and resources.
- Provide general insight into the uncertainties in the evaluated energy savings estimates for NYSERDA's **New York Energy $mart**^{SM} programs based on the results of past evaluations.
- Permit future re-evaluation of risk to determine if key areas of uncertainty have been reduced.

The underlying, oft-cited reasons for using risk analysis are to identify, quantify, and manage risk. This is done through risk analysis methods that quantify full, probabilistic distributions of values, uncertainty, and the underlying drivers of such uncertainty. NYSERDA characterized three types of uncertainty from prior evaluations: methodological, physical, and attributional uncertainties.

- Methodological uncertainties address the accuracy and appropriateness of the savings algorithms or source of the deemed savings.
- Physical uncertainties are associated with the measures used as inputs into algorithms or for determining deemed savings (*e.g.*, hours-of-use, pre-/post-retrofit efficiency, etc.).
- Attribution uncertainties relate to how much of the realized savings can be claimed by the program (including uncertainties within the net-to-gross ratios and their components).

Since this was NYSERDA's first time applying probabilistic risk analysis in the context of impact evaluation, the following decisions were made by the evaluation team to ensure that this initial risk analysis project was simple, manageable, and provided useful results:

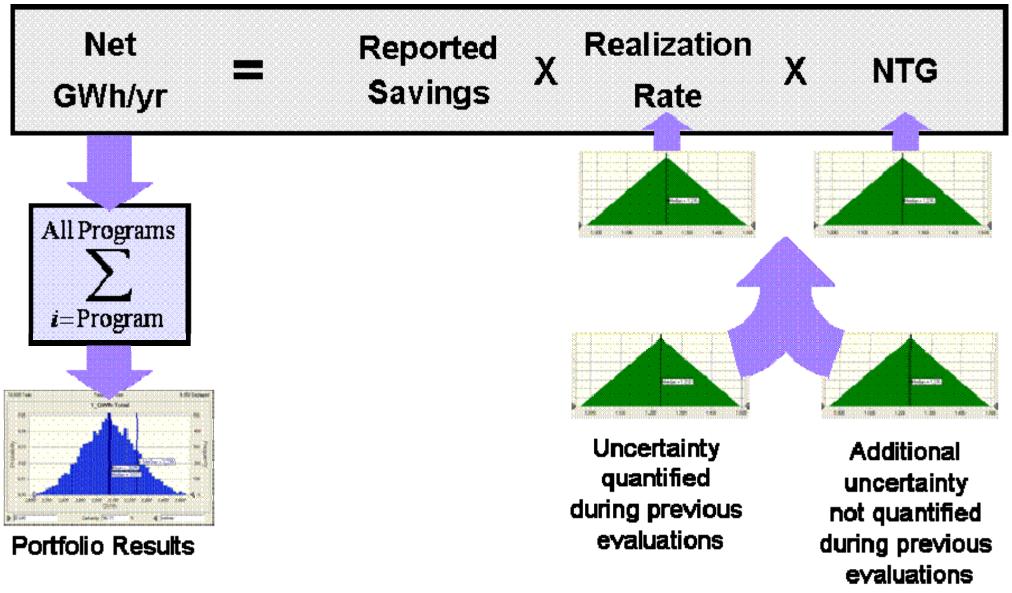
- The input data that determined the distribution of risk were based on past impact evaluations, as well as the expert professional judgment of the Megdal & Associates' (M&A) Team. Since the M&A Team was reviewing previous evaluations as part of their overall effort, this allowed them to quickly and efficiently develop input distributions for risk.
- The analysis focused on the twelve largest NYSERDA programs that had detailed previous impact evaluations. For each program, uncertainty around the prior measurement and verification (M&V) gross savings realization rates (RR) and around the net-to-gross ratio (NTG) were considered. Therefore, the entire risk model had twenty-four input distributions—one for RR and one for NTG for each of the twelve programs.
- The NYSERDA risk analysis addressed only electric energy (GWh) savings. Demand reductions (MW) achieved by one load management program were also considered.
- Since the primary objective of the risk analysis was to aid the allocation of evaluation resources, the project did not consider methods of optimizing NYSERDA's energy efficiency portfolio.
- Since previous impact evaluations were the primary sources of input data, the team conducted its analysis at the program and portfolio levels. Risk analysis below the program level (*e.g.,* measure level) was beyond the scope and would not have been possible given data availability and budget. Further consideration of measure-level risk analysis for RR estimates may occur in the future if budgets permit given other evaluation priorities.

Methodological Overview

Over 35 reports from past evaluations were reviewed to determine what type of information would be readily available on uncertainty surrounding the previous estimates. Then, the model was developed and implemented. The equation shown in Figure 1 is the simple calculation of net savings: reported savings x RR x NTG ratio = net savings. For each RR and NTG ratio, uncertainties were identified from two sources: (1) those explicitly discussed in the prior evaluation reports, and (2) those determined through qualitative reliability assessment by evaluation experts. The evaluation reports reviewed typically only provided a quantitative estimate of uncertainty resulting from sampling precision. Additional uncertainty, beyond sampling error, was determined by examining drivers such as: research methods, use of construct validity testing, time elapsed since the last evaluation, and consistency of results over time. These two sources of uncertainty were then blended together to develop a single input distribution that represents the uncertainty around each of the two independent variables—RR and NTG. The specific steps surrounding the quantification of uncertainty are described in more detail in the next section of this paper. A *Monte Carlo* simulation used the uncertainty distributions of RR and NTG to generate the distribution of uncertainty in the overall evaluation savings estimates at the portfolio level. This approach directly incorporates the program-reported savings in the portfolio uncertainty analysis. So the model itself already addresses the effect of program size. This simplifies the analysis by avoiding any weighting or other adjustments.

Monte Carlo simulation uses a distribution of values (as opposed to a simple scalar value) as the inputs to the independent variables of a mathematical model. Input values are randomly selected per these distributions to generate another distribution of the resulting values for the dependent variable(s) in question. In addition to simply generating distributions of the dependent variable, *Monte Carlo* simulation software calculates the sensitivity (*i.e.*, contribution to variance) of the result to each of the input variables, thus indicating which variables contribute most to the uncertainties (*i.e.*, quantifying the largest drivers of risk). Therefore, the methodology used to assign the uncertainty distributions to the input variables is critical in developing an effective risk model.

Figure 1. Method for Calculating Input Distributions



Quantifying Uncertainty

All of the input distributions for this risk analysis were developed by the M&A Team's review of past impact evaluation reports and expert judgment to convert the qualitative reliability assessment parameters to relative risk profiles. To ensure consistency across all programs examined, a single expert assigned the qualitative risk values for RR and another expert assigned risk values for NTG. Other possible alternative approaches would have been for a single expert to assign the risk values for *both* the RR and NTG, or having multiple experts create the risk assignments. NYSERDA balanced the need for consistency and project cost-efficiency by splitting the work into two halves so there was no inter-rater induced variation in the risk assessment between programs while meeting NYSERDA's desire to accomplish the project in a reasonably short time frame.

Gross Savings Uncertainty

Gross savings were estimated in previous impact evaluations by adjusting the original NYSERDA program-reported savings by a realization rate. The previous evaluations calculated and reported an 80% confidence interval (CI) around the RR. This confidence interval served as the base for the uncertainty analysis around the RR. However, since the CI captures only the uncertainties based on sampling error (sampling precision), additional sources of uncertainty will only increase the actual level of uncertainty associated with the RR and gross savings. Thus, a more inclusive estimate of overall uncertainty was modeled by adjusting the CI to take into account other contributing factors.

Categories of additional uncertainty were identified in two groups: program uncertainties (PU) and evaluation uncertainties (EU). Each additional category of uncertainty within these two groups was assigned a weighting that estimates its potential maximum effect on the confidence interval. The weightings were based on the expert judgment of the evaluation team. This methodology provides a consistent and transparent framework for quantifying a subjective value. Table 1 and Table 2 show the categories of uncertainty and the weighting assigned to each within the groups of program and evaluation uncertainty, respectively.[2]

---

[2] One may notice that the "Potential Effect on CI" column adds up to one in Table 1 but not in Table 2. This column reflects the actual range of variation in the RR that is associated with each category, rather than a percentage of the total change in the RR related to the category, and thus would not be expected to total to one. In Table 1, all categories are not applicable to all programs and the column total for a specific program would probably not be equal to one.

Table 1. Categories of Program Uncertainty for Realization Rate

Categories of Program Uncertainty	Potential Effect on CI (Weight)	Description
Savings method	0.20	Method of estimating savings (deemed, custom with standardized software, etc.)
Known installation rates	0.20	Whether installations are recorded and/or verified as part of program delivery
Program changes	0.20	Major changes in program procedures
Baseline	0.10	Uncertainty associated with the relative difficulty of defining the baseline
Range of Measures	0.10	Reflecting the range of types of measures (from single measures to complex analyses with many components)
Program tracking	0.10	Comprehensiveness of program tracking mechanism
Timing/lag in installation	0.05	Length of lead time between the audit or analysis and installation of measures
Retention	0.05	Whether measures are likely to remain in place or be removed/replaced

Table 2. Categories of Evaluation Uncertainty for Realization Rate

Categories of Evaluation Uncertainty	Potential Effect on CI (Weight)	Description
RR variation over time	0.20	The degree to which the RR varies from one evaluation period to another
Sample size/design	0.10	Uncertainty added due to sample size and/or design issues
Research method	0.10	Overall uncertainty from research method used
Years since last evaluation	0.05	Used 2008 as base to measure "years from" given report dates used for "when evaluated"
Verification of evaluation method/estimate	0.02	Alternative method used for verification of estimate

Each category of PU and EU was assigned a score for each program on a scale from zero-to-three. A score of zero meant that the specific category did not apply and no additional uncertainty would be added to the base confidence interval. A score of three implied that the specific program had a high level of uncertainty in that particular category. Based on this scoring and weighting methodology, an adjusted CI was calculated by taking the sum of the products of each potential effect with the scoring for the PUs. The same method was used for the EUs, providing two different methods of estimating the uncertainty for each program. At the program level, an additional step was needed to account for the sampling error associated with the previous evaluations. Consequently, the adjusted CI calculated through the exercise described above was added to the original CI from the previous evaluations, where available.

The final adjusted CI for RR is the greater of the adjusted PU CI and the adjusted EU CI. The final input distribution was built with the expected value of the RR equal to the RR, a lower bound equal to the RR minus the adjusted confidence interval, and an upper bound equal to the RR plus the adjusted confidence interval. The final lower and upper bounds also included an assessment of whether the CI distribution would be equally spread around the expected value or skewed. Skew (*e.g.*, potential directional bias) was represented by dividing the CI into different proportions on each side of the expected value.

Net Savings Uncertainty

The general methodology for quantifying the input distributions around the net-to-gross ratios is similar to that of calculating the input distributions for RR. However, the NTG uncertainty is slightly less complex because of the lack of distinction between program and evaluation uncertainties. Twelve categories of uncertainty were developed and given a relative weighting of importance as shown in Table 3. The weighting is relative to an average distribution of the CI. Each program evaluation was scored according to each category. The RR and NTG uncertainties needed to be calculated separately to ensure the detailed analysis for each of their evaluation components. Yet, the analysis also needed to ensure that the adjustments to the CI from the RR and NTG assessments were scaled relative to one another to ensure the appropriate assignment of uncertainty to each part of the analyses. To make this relative scale element explicit, the NTG model used a conversion factor to represent relative scale. This conversion factor was used to convert weighting to increased confidence interval as the proxy of uncertainty bands. For this analysis the conversion factor was 50%. For each program, the adjusted NTG confidence interval was calculated by multiplying the score by the weight and the conversion factor for each category of uncertainty. This adjusted confidence interval was then used to generate a triangle distribution for each of the programs. The final lower and upper bounds also included an assessment of whether or not the CI distribution would be skewed (*e.g*., potential bias), and any skew was represented by dividing the CI into different proportions on each side of the expected value.

Table 3. Categories of Net-to-Gross Uncertainty

Categories of NTG Uncertainty	Weight	Description
Evaluation relevance for ‘06-‘07	0.18	Degree of program change since the latest NTG was estimated
Potential bias	0.18	Degree to which review pointed to a potential bias in the estimate
Market-based	0.15	Degree of market-based components
Method & research design	0.12	Overall uncertainty from method & research design used
Construct validity test	0.10	Alternative methods or tests to increase construct validity
Years since last evaluation	0.06	Used 2008 as base to measure "years from" given report dates used for "when evaluated" with maximum years set at seven
Free-ridership inconsistency	0.05	Degree to which the method handled inconsistent responses on free-ridership
Evaluation coverage for most of the program savings	0.05	The degree to which the evaluation covered the program savings (less coverage = more uncertainty)
Sample design	0.05	Uncertainty added due to sample design issues
Spillover inconsistency	0.03	How well did the method handle inconsistent responses concerning spillover?
NTG instability over time	0.02	The degree to which the NTG varies with time
Triangulation	0.01	Alternative actors used for triangulation of estimate

Scope and Results

Table 4 shows the 12 NYSERDA programs included in the risk analysis. The risk analysis focuses primarily on uncertainty surrounding the biggest contributors to the portfolio-level electricity savings. However, the evaluation team also examined risk surrounding the Peak Load Reduction Program, which is the largest contributor to the portfolio's demand (MW) reduction achievements.

Table 4. NYSERDA Programs Included in Risk Analysis

Program Name	Abbreviation	Unit of Savings	Net Savings Reported by NYSERDA (Q1 2007)
Commercial/Industrial Performance Program	CIPP	GWh	854
Technical Assistance FlexTech & Energy Audits	TA	GWh	738
ENERGY STAR Products	ESProducts	GWh	647
New Construction Program	NCP	GWh	277
Distributed Generation/Combined Heat and Power	DG/CHP	GWh	101
Smart Equipment Choices	SEC	GWh	83
Small Commercial Lighting Program	SCLP	GWh	38
Home Performance with ENERGY STAR	HPwES	GWh	15
New York ENERGY STAR Labeled Homes	NYESLH	GWh	10
Premium-Efficiency Motors Program	PEM	GWh	9
Commercial HVAC Program	CIHVAC	GWh	7
Peak Load Reduction Program	PLRP	MW	615
Total GWh			**2,779**

Electricity Savings Risk Analysis Results

A summary of the results from the simulations is presented in Table 5. Column A shows the mean net electricity savings as adjusted by the risk analysis results. Column B shows the minimum net adjusted savings. Column C shows the range of possible electricity savings around the mean, and Column D divides the range by the mean savings. Column C shows that the range of savings is greatest for the largest programs as would be expected in any analysis that properly incorporates the size of the program in relation to the portfolio.

Another important finding is the relative uncertainty shown in Column D. This allows evaluation planners to "see" where using additional resources or better research design could most improve savings estimates. Another finding and benefit from this analysis can be found through the observation that the relative uncertainty for the larger programs versus the smaller programs is lower on average. This demonstrates NYSERDA's effectiveness in devoting evaluation resources to those programs that contribute the greatest savings to the portfolio to reduce risk in those savings estimates and, therefore, to reduce risk in the evaluated savings estimates of the portfolio.

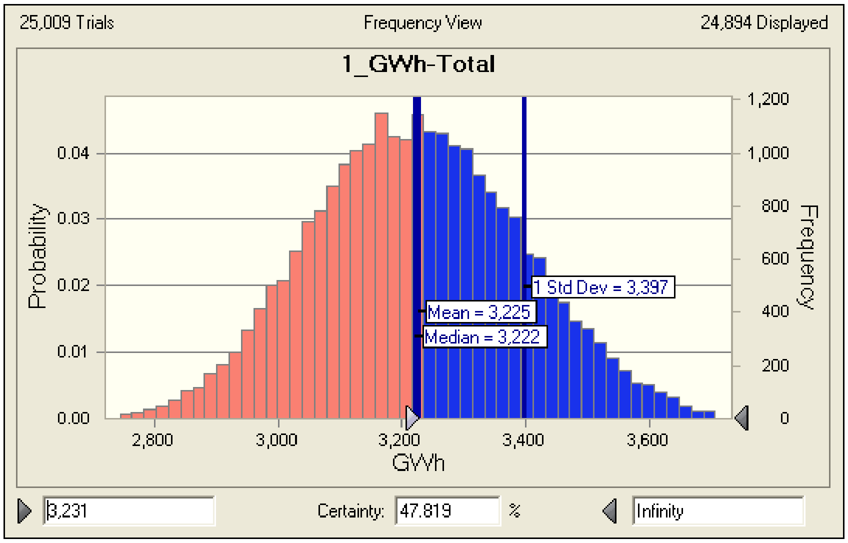
Comparing Column A of Table 5 to the last column of Table 4 indicates that, for the most part, the mean net adjusted electricity savings values from the risk analysis are higher than the net savings being claimed by NYSERDA. Furthermore, many of the ranges are somewhat skewed toward the high end. Thus, one might postulate that steps to further reduce uncertainty on these programs could lead to an increase in savings. However, decreases are also possible as indicated by the minimum values shown in Table 5.

Table 5. Summary of Uncertainty by Program

Program	A	B	C	D
	Mean Savings (GWh)	Minimum Savings (GWh)	Range of Savings (GWh)	Relative Uncertainty (Range/Savings) (C/A)
CIPP	1,048	723	703	67%
TA	860	589	598	70%
ESProducts	647	483	347	54%
NCP	391	225	355	91%
DG/CHP	111	62	116	105%
SCLP	75	37	96	128%
SEC	45	20	57	126%
HPwES	17	10	15	88%
NYESLH	14	8	15	109%
PEM	6	2	8	130%
CIHVAC	1	0	2	155%

Figure 2 shows the distribution of the portfolio's cumulative adjusted net GWh savings. The distribution is symmetric and centered around the expected value. This is primarily driven by the fact that none of the evaluations of the large programs had evidence of bias.

Figure 2. Distribution of Portfolio's Cumulative Adjusted Net GWh Savings



3,231

Certainty: 47.819 %

Infinity

Figure 3 shows which inputs contribute to portfolio level variance. Often one or two input variables dominate the sensitivity charts. In this case, however, six uncertainties contribute most (93%) of the portfolio's variance. Uncertainty in NTG for the CIPP contributes most of the risk (30.4%), but the gross GWh uncertainty for CIPP appears 5th on the list (12%). From a program perspective, the CIPP (42.4%) and TA Program (31.2%) contribute most to the overall portfolio risk.

Figure 3. Portfolio Sensitivity Chart

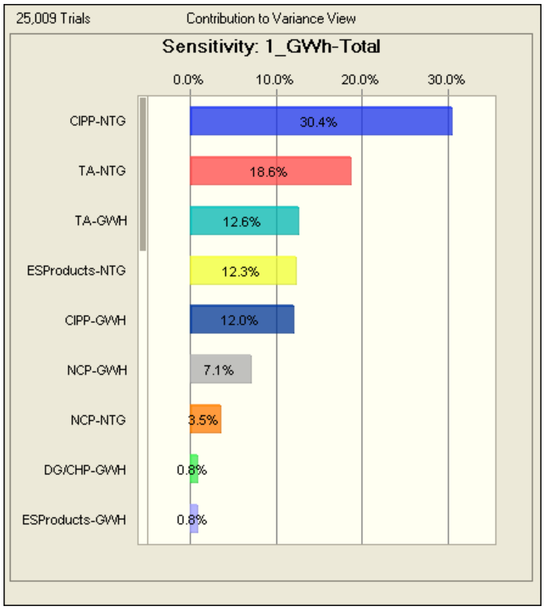


Figure 4 shows which programs have the most uncertainty. As expected, larger programs have larger uncertainties and contribute most to the portfolio uncertainty. Nevertheless, there were programs where the uncertainty range was greater than the expected value. However, most were relatively small energy savers.

Figure 4. Median and Range of Cumulative Adjusted Net GWh Savings by Program

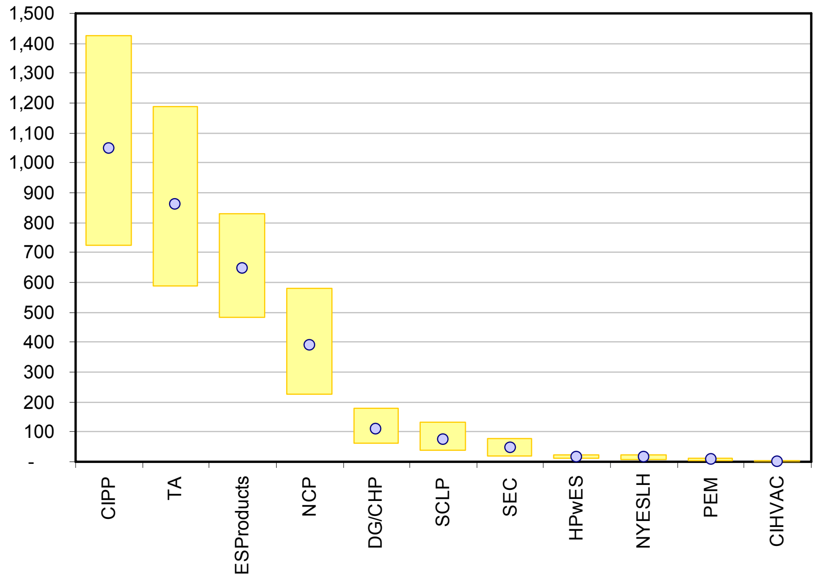
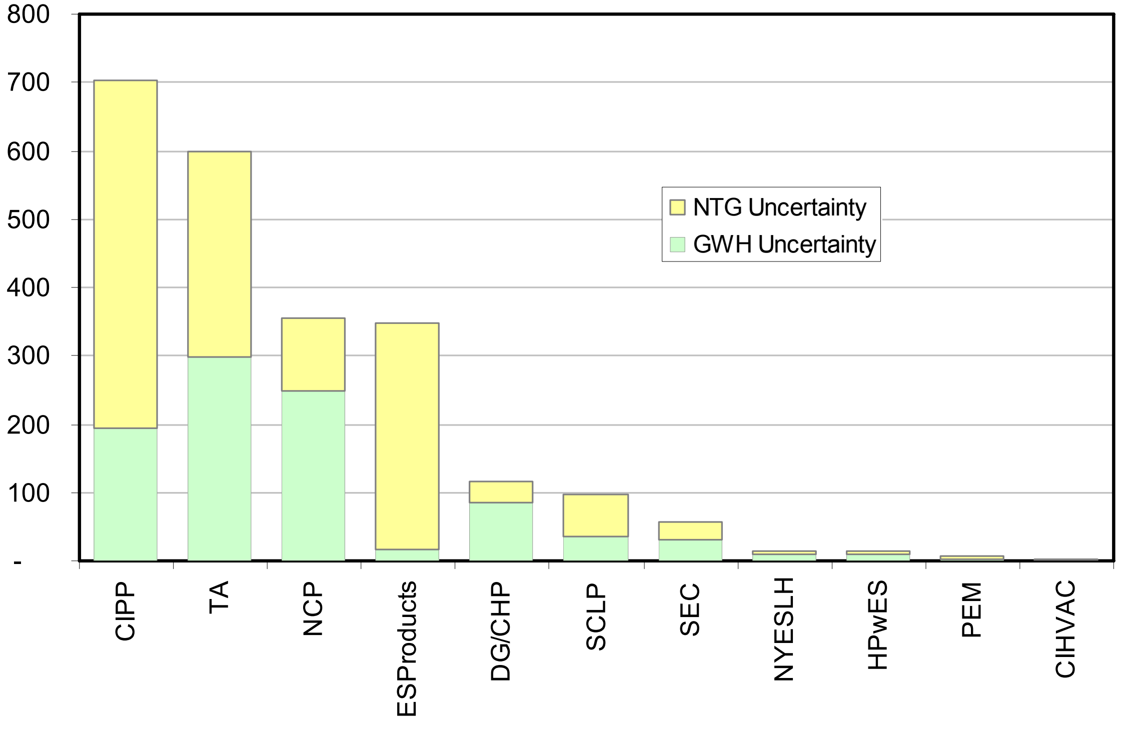


Figure 4 alone would support one's intuition for allocating evaluation resources to the programs with the largest energy savings. However, a non-intuitive nuance becomes clear from analyzing Figure 5 which shows which components of uncertainty (either RR or NTG) are most dominant. Some key observations include the following:

- Although CIPP has the largest uncertainty, most of that uncertainty comes from the NTG component of the savings estimate. This is reasonable since the CIPP implementation requires detailed M&V. Therefore, adding copious evaluation resources to reduce the uncertainty around the RR may not be necessary or economical. More emphasis should be placed on the attribution issues instead.
- The second greatest uncertainty is with the TA ex-post estimates. Contrary to the CIPP, this program shows uncertainty equally split between gross savings estimates and the net-to-gross ratio. The greater proportion of uncertainty in RR for this program compared to CIPP is due to uncertainty associated with not knowing who and what measures were adopted (since this is a technical audit effort) and also with the savings estimates themselves.
- Most of the uncertainty in NCP is related to the gross savings. One of the primary contributors to this uncertainty is related to the need for additional baseline work.
- Most of the uncertainty in ESProducts is driven by attribution. Uncertainty around the gross savings is very low given the recent evaluations of residential lighting and appliance equipment savings in New England and elsewhere. The attribution issues may indicate the need for further market analyses.
- The DG/CHP Program shows greater uncertainty in the gross savings than NTG. The largest contributor to uncertainty was within the data tracking efforts.

Figure 5. Components of Uncertainty Range in GWh by Program



Demand Reduction Risk Analysis Results

The risk analysis produced similar results for the Peak Load Reduction Program's demand reductions. Similar to Figure 2 for the net GWh savings, the distribution of MW reductions is symmetric around the expected value. This is directly caused by the fact that the input distributions in RR and NTG were also both symmetric and assessed as unbiased. Regarding the components of uncertainty, the PLRP

NTG uncertainty (71%) was greater than the RR uncertainty (29%) as the program has required field verification for most of its projects.

Application and Lessons Learned

NYSERDA's first risk analysis provided evidence that early evaluations, which allocated evaluation resources to the largest energy-saving programs, were appropriately directed at reducing portfolio level risks. To NYSERDA's evaluation staff and managers, the high level results of the risk analysis largely confirmed what their experience evaluating these programs suggested in terms of the programs where the risk may be the greatest. However, this work did shed new light on the specific components of the program-level savings estimates that were driving the risk. The risk analysis results presented in this paper were used by NYSERDA as one of the inputs to guide allocation of the impact evaluation budget. However, additional factors beyond the risk analysis were also considered in selecting impact evaluation projects. One of the most important of these is the cost of reducing the identified uncertainty. There are some causes of uncertainty that cannot be easily addressed by modifying the evaluation methodology, research design, or sampling. One obvious example is the inherent uncertainty in market-based analyses as was employed for the ESProducts Program. Other project selection criteria besides budget and value constraints may include program modification needs, policy needs, and ability to leverage other evaluation projects to gather data.

The risk analysis has lent additional confidence to decisions related to impact evaluation resource allocation. Based on the risk analysis results and other considerations, NYSERDA has enhanced and modified its approach to evaluating gross and net savings in this evaluation cycle. Instead of conducting individual program-level evaluations of RR and NTG on a regular one-to-two year time frame, NYSERDA is conducting intensive, in-depth site visits and enhanced net-to-gross evaluation on approximately 30 of the largest energy-saving projects across the portfolio. Reducing uncertainty on these largest projects will result in a decrease in overall portfolio uncertainty. NYSERDA is also incorporating new methods to further cross check and validate NTG results. In this area, NYSERDA will employ an "enhanced self-report" approach for its "largest energy-savers evaluation," not just relying on the standard battery of NTG questions, but also asking participating customers for documentation of their decision-making, financing, and purchase policies regarding the energy-efficient equipment they replaced as part of the program. NYSERDA is also adding new questions to the NTG battery for other program evaluations to test construct validity.

Risk analysis is often said to create ancillary benefits simply by having organizations go through the process. The identification and quantification of risk is the necessary precursor to being able to manage it. Explicitly reviewing past evaluations for reliability issues and then developing weighting and scoring for each evaluation's qualitative reliability assessment was enlightening, and NYSERDA plans to go back and revisit the risk analysis for the next post evaluation period. In this way, a continuing improvement process can be implemented. Using the risk analysis for systematic improvement of impact evaluations is an innovative strategy that should assist NYSERDA in moving toward ever more robust evaluations within its budget constraints.

NYSERDA's positive first experience and the ongoing potential for this approach provide a solid foundation to recommend this type of analysis to other evaluators and organizations. It would also be quite helpful to have others study and develop further the methods of translating qualitative reliability assessments into quantitative views of uncertainty. Opening a dialogue across the efficiency evaluation field in this research area could help increase reliability and lead to better evaluations.

References

Hall, Nick, Pete Jacobs, and Steve Kromer. 2006. *Improving the Reliability of Energy Efficiency Portfolio and Program Evaluation Via a Risk Assessment Approach*. Prepared for the Energy Division of the California Public Utilities Commission, 2006.

NYSERDA. 2007. Risk Analysis for NYSERDA's System Benefit Charge-Funded Impact Evaluation Estimates. Prepared for NYSERDA's Energy Analysis Group by Rational Energy Network, West Hill Energy & Computing, and Megdal & Associates. November 2007.

Ridge, Richard, Steve Kromer, Steven Meyers, Doug Mahone, Jay Luo, Valerie Richardson, Rafael Friedmann, and Nick Hall. 2007. "Energy Efficiency Portfolio Risk Management: A Systematic, Data-Driven Approach for Timely Interventions to Maximize Results", *Proceedings of the 2007 International Energy Program Evaluation Conference*. August: Chicago, Ill., pp. 1006-1077.

Violette, Daniel and Kevin Cooney. 2003. *Findings and Report: Retrospective Assessment of the Northwest Energy Efficiency Alliance*, Prepared for: Northwest Energy Efficiency Alliance Ad Hoc Retrospective Committee, by Summit Blue Consulting, December 8, 2003.